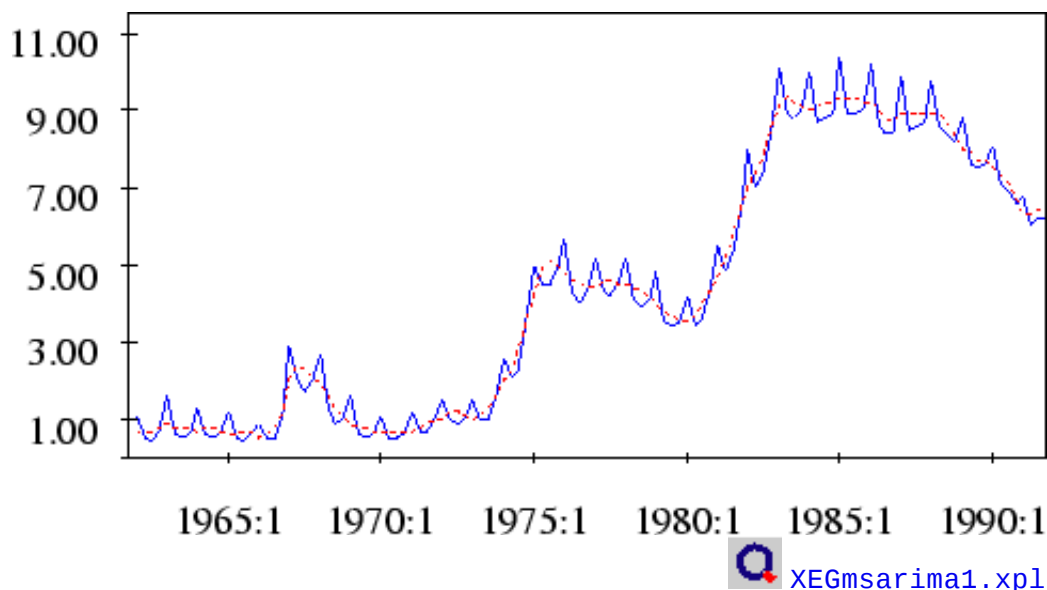


## 5.2 MODELING SEASONAL TIME SERIES

### 5.2.1 Seasonal ARIMA Models





Before one can specify a model for a given data set, one must have an initial guess about the data generation process. The first step is always to plot the time series. In most cases such a plot gives first answers to questions like: "Is the time series under consideration stationary?" or "Do the time series show a seasonal pattern?"

Figure 5.1 displays the quarterly unemployment rate  $u_t$  for Germany (West) from the first quarter of 1962 to the fourth quarter of 1991. The data are published by the OECD ([Franses; 1998](#), Table DA.10).



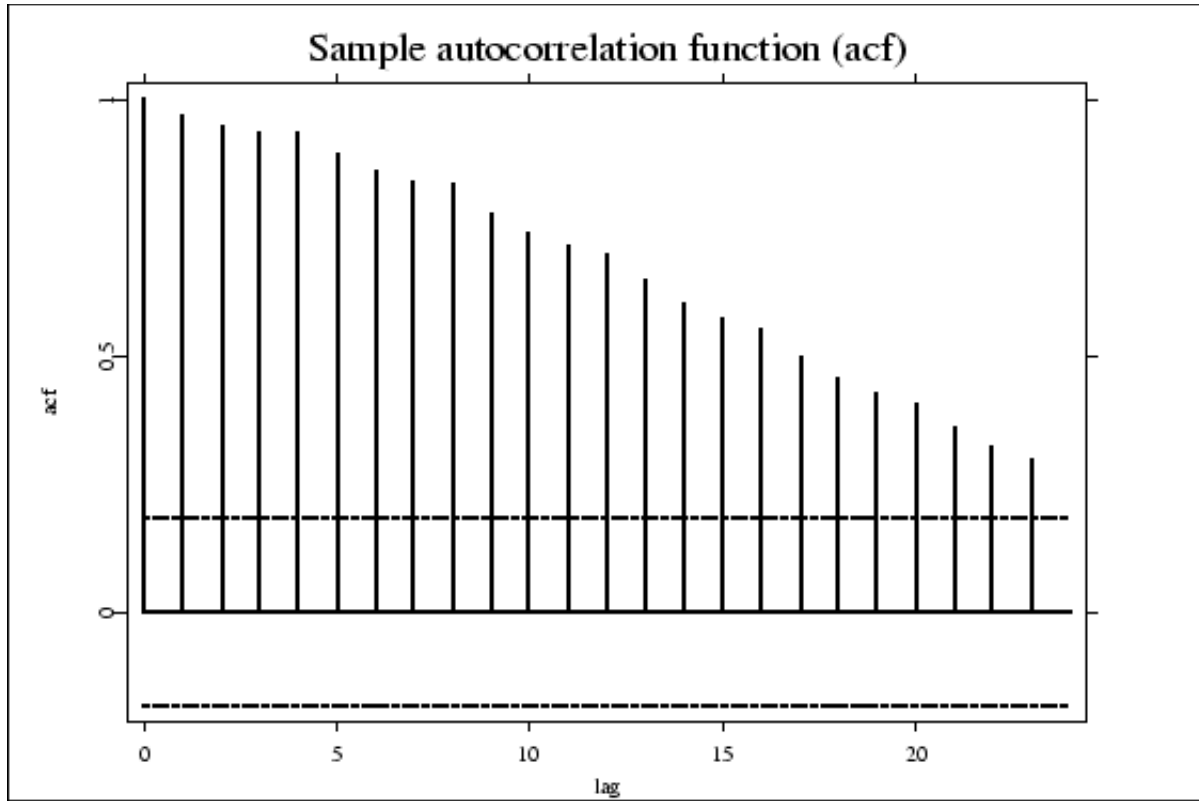
**Figure 5.1:** Quarterly unemployment rate for Germany (West) from 1962:1 to 1991:4. The original series ( $u_t$ ) is given by the solid blue line and the seasonally adjusted series is given by the dashed red line.

The solid line represents the original series  $u_t$  and the dashed line shows the seasonally adjusted series. It is easy to see, that this quarterly time series possesses a distinct seasonal pattern with spikes recurring always in the first quarter of the year.

After the inspection of the plot, one can use the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) to specify the order of the ARMA part (see  [acf](#) ,  [pacf](#) ,  [acfplot](#) and  [pacfplot](#) ). Another convenient tool for first stage model specification is the extended autocorrelation function (EACF), because the EACF does not require that the time series under consideration is stationary and it allows a simultaneous specification of the autoregressive and moving

average order. Unfortunately, the EACF can not be applied to series that show a seasonal pattern. However, we will present the EACF later in Section 5.4.5, where we use it for checking the residuals resulting from the fitted models.

Figures 5.2, 5.3 and 5.4 display the sample ACF of three different transformations of the unemployment rate ( $u_t$ ) for Germany.



 [XEGmsarima2.xpl](#)

**Figure 5.2:** Sample ACF of the unemployment rate for Germany (West) ( $u_t$ ) from 1962:1 to 1991:1.

Using the difference--or backshift--operator  $L$ , these kinds of transformations of the unemployment rate can be written compactly as

$$\Delta^d \Delta_s^D u_t = (1 - L)^d (1 - L^s)^D u_t,$$

where  $L^s$  operates as  $L^s u_t = u_{t-s}$  and  $s$  denotes the seasonal period.  $\Delta^d$  and  $\Delta_s^D$  stand for nonseasonal and seasonal differencing. The superscripts  $d$  and  $D$  indicate that, in general, the differencing may be applied  $d$  and  $D$  times.

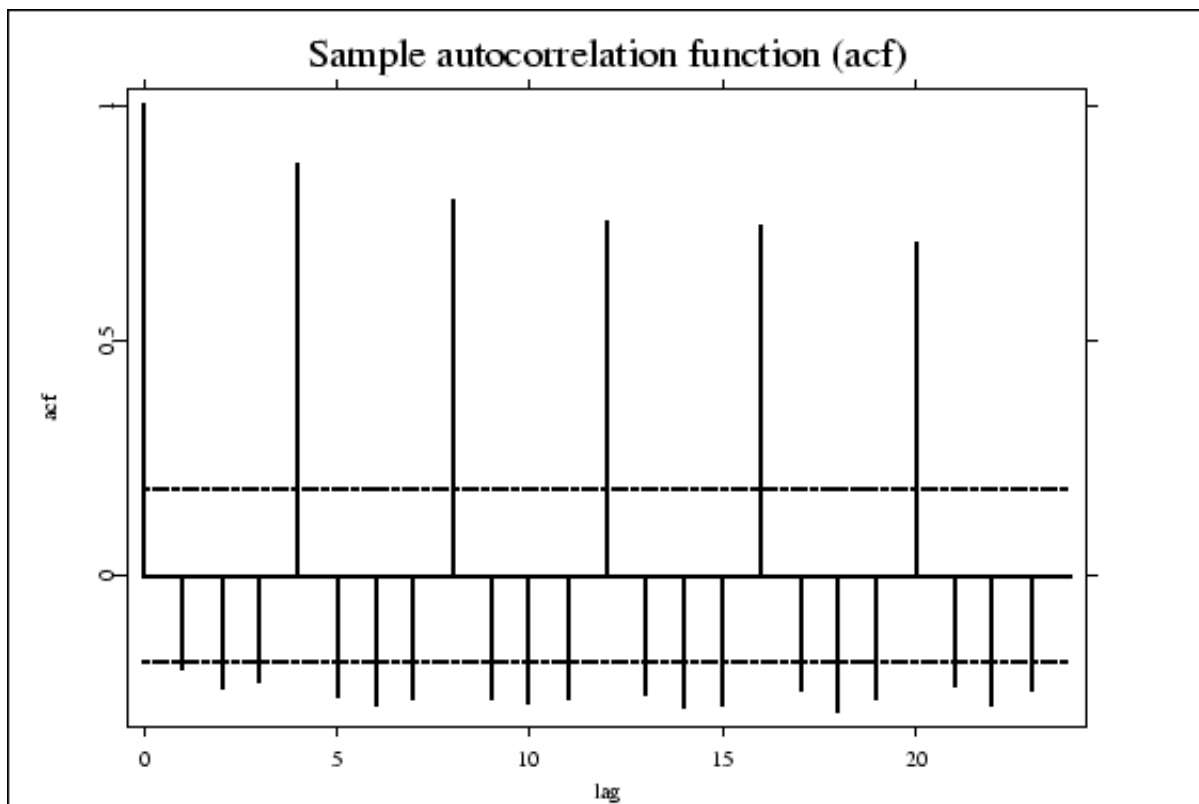
Figure 5.2 shows the sample ACF of the original data of the unemployment rate ( $u_t$ ). The fact, that the

time series is neither subjected to nonseasonal nor to seasonal differencing, implies that  $d = D = 0$ . Furthermore, we set  $s = 4$ , since the unemployment rate is recorded quarterly. The sample ACF of the unemployment rate declines very slowly, i.e. that this time series is clearly nonstationary. But it is difficult to isolate any seasonal pattern as all autocorrelations are dominated by the effect of the nonseasonal unit root.

Figure 5.3 displays the sample ACF of the first differences of the unemployment rate ( $\Delta u_t$ ) with

$$\Delta u_t = u_t - u_{t-1}.$$

Since this transformation is aimed at eliminating only the nonseasonal unit root, we set  $d = 1$  and  $D = 0$ . Again, we set  $s = 4$  because of the frequency of the time series under consideration.



[XEGmsarima3.xpl](#)

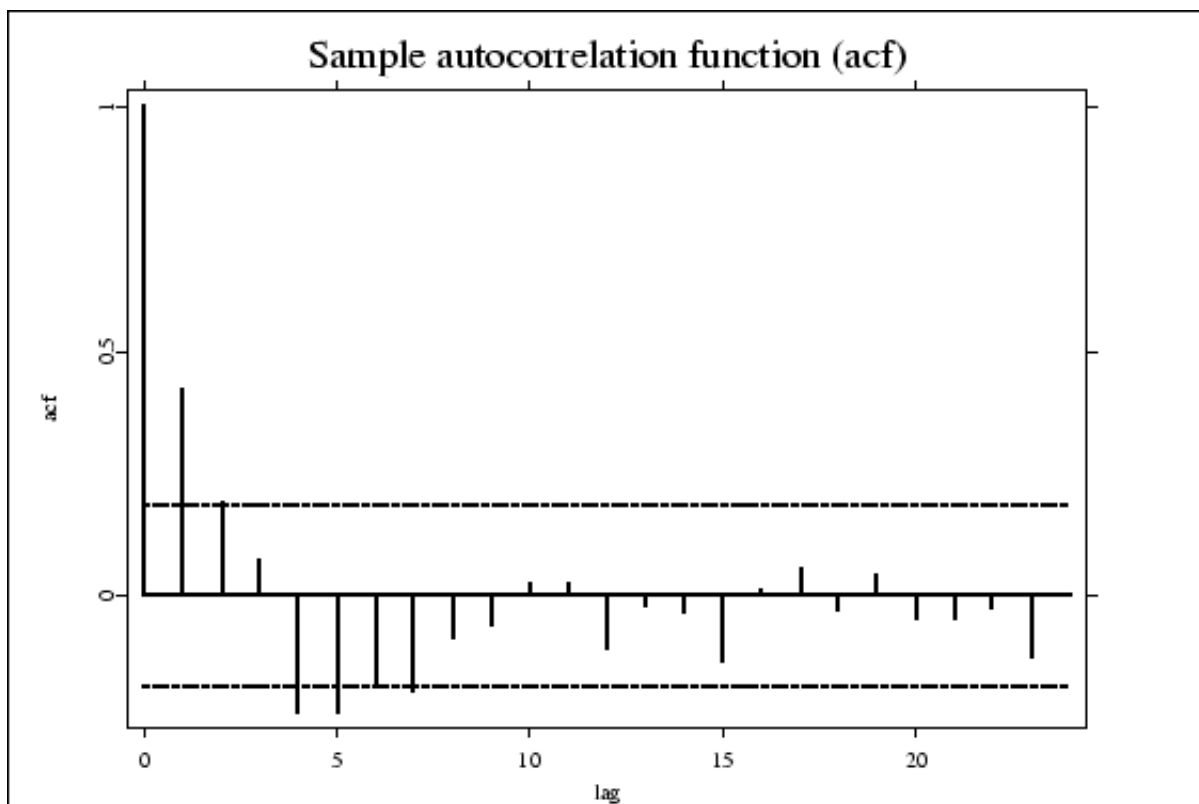
**Figure 5.3:** Sample ACF of the first differences of the unemployment rate ( $\Delta u_t$ ) for Germany (West) from 1962:1 to 1991:1.


Taking the first differences produces a very clear pattern in the sample ACF. There are very large positive autocorrelations at the seasonal frequencies (lag 4, 8, 12, etc.), flanked by negative autocorrelations at the 'satellites', which are the autocorrelations right before and after the seasonal lags. The slow decline of the seasonal autocorrelations indicates seasonal instationarity. Analogous to the analysis of nonseasonal

nonstationarity, this may be dealt by seasonal differencing; i.e. by applying the  $\Delta_4 = (1 - L^4)$  operator in conjunction with the usual lag operator  $\Delta = (1 - L)$  (Mills: 1990, Chapter 10).

Eventually, Figure 5.4 displays the sample ACF of the unemployment rate that was subjected to the final transformation

$$\begin{aligned}\Delta\Delta_4 u_t &= (1 - L)(1 - L^4)u_t \\ &= (u_t - u_{t-4}) - (u_{t-1} - u_{t-5}) .\end{aligned}$$



 [XEGmsarima4.xpl](#)

**Figure 5.4:** Sample ACF of the seasonally differenced first differences of the unemployment rate (  $\Delta\Delta_4 u_t$  ) for Germany (West) from 1962:1 to 1991:1.

Since this transformation is used to remove both the nonseasonal and the seasonal unit root, we set  $d = D = 1$ . What the transformation  $\Delta\Delta_4$  finally does is seasonally differencing the first differences of the unemployment rate. By means of this transformation we obtain a stationary time series that can be modeled by fitting an appropriate ARMA model.

After this illustrative introduction, we can now switch to theoretical considerations. As we already saw in practice, a seasonal model for the time series  $\{x_t\}_{t=1}^T$  may take the following form

$$\Delta^d \Delta_s^D x_t = \frac{\Theta(L)}{\Phi(L)} a_t, \quad (5.1)$$

where  $\Delta^d = (1 - L)^d$  and  $\Delta_s^D = (1 - L^s)^D$  indicate nonseasonal and seasonal differencing and  $s$  gives the season.  $a_t$  represents a white noise innovation.  $\Phi(L)$  and  $\Theta(L)$  are the usual AR and MA lag operator polynomials for ARMA models

$$\Phi(L) \equiv 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

and

$$\Theta(L) \equiv 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q.$$

Since the  $\Phi(L)$  and  $\Theta(L)$  must account for seasonal autocorrelation, at least one of them must be of minimum order  $s$ . This means that the identification of models of the form (5.1) can lead to a large number of parameters that have to be estimated and to a model specification that is rather difficult to interpret.

## 5.2.2 Multiplicative SARIMA Models

[Box and Jenkins \(1976\)](#) developed an argument for using a restricted version of equation (5.1), that should be adequate to fit many seasonal time series. Starting point for their approach was the fact, that in seasonal data there are two time intervals of importance. Suppose, that we still deal with a quarterly series, we expect the following to occur ([Mills: 1990](#), Chapter 10):

- a seasonal relationship between observations for the same quarters in successive years, and
- a relationship between observations for successive quarters in a particular year.

Referring to Figure 5.1 that displays the quarterly unemployment rate for Germany, it is obvious that the seasonal effect implies that an observation in the first quarter of a given year is related to the observations of the first quarter for previous years. We can model this feature by means of a *seasonal model*

$$\Phi_s(L)\Delta_s^D x_t = \Theta_s(L)v_t . \quad (5.2)$$

$\Phi_s(L)$  and  $\Theta_s(L)$  stand for a seasonal AR polynomial of order  $p$  and a seasonal MA polynomial of order  $q$  respectively:

$$\Phi_s(L) = 1 - \phi_{s,1}L^s - \phi_{s,2}L^{2s} - \dots - \phi_{s,P}L^{Ps}$$

and

$$\Theta_s(L) = 1 + \theta_{s,1}L^s + \theta_{s,2}L^{2s} + \dots + \theta_{s,Q}L^{Qs} ,$$

which satisfy the standard stationarity and invertibility conditions.  $v_t$  denotes the error series. The characteristics of this process are explained below.

It is obvious that the above given seasonal model (5.2) is simply a special case of the usual ARIMA model, since the autoregressive and moving average relationship is modeled for observations of the same seasonal time interval in different years. Using equation (5.2) relationships between observations for the same quarters in successive years can be modeled.

Furthermore, we assume a relationship between the observations for successive quarters of a year, i.e. that the corresponding error series (  $v_t, v_{t-1}, v_{t-2}$ , etc.) may be autocorrelated. These autocorrelations may be represented by a *nonseasonal model*

$$\Phi(L)\Delta^d v_t = \Theta(L)a_t . \quad (5.3)$$

$v_t$  is ARIMA  $(p, d, q)$  with  $a_t$  representing a process of innovations (white noise process).

Substituting (5.3) into (5.2) yields the *general multiplicative seasonal model*

$$\Phi(L)\Phi_s(L)\Delta^d\Delta_s^D x_t = \delta + \Theta(L)\Theta_s(L)a_t . \quad (5.4)$$

In equation (5.4) we additionally include the constant term  $\delta$  in order to allow for a deterministic trend in the model ([Shumway and Stoffer; 2000](#)). In the following we use the short-hand notation SARIMA (

$p, d, q) \times (s, P, D, Q)$  to characterize a multiplicative seasonal ARIMA model like (5.4).

### 5.2.3 The Expanded Model

Before to start with the issues of identification and estimation of a multiplicative SARIMA model a short example may be helpful, that sheds some light on the connection between a multiplicative SARIMA ( $p, d, q) \times (s, P, D, Q)$  and a simple ARMA ( $p, q$ ) model and reveals that the SARIMA methodology leads to parsimonious models.

Polynomials in the lag operator are algebraically similar to simple polynomials  $ax + bx^2$ . So it is possible to calculate the product of two lag polynomials (Hamilton, 1994, Chapter 2).

Given that fact, every multiplicative SARIMA model can be telescoped out into an ordinary ARMA( $p, q$ ) model in the variable

$$y_t \stackrel{\text{def}}{=} \Delta_s^D \Delta^d x_t .$$

For example, let us assume that the series  $\{x_t\}_{t=1}^T$  follows a SARIMA( 0, 1, 1)  $\times$  (12, 0, 1, 1) process.

In that case, we have

$$(1 - L^{12})(1 - L)x_t = (1 + \theta_1 L)(1 + \theta_{s,1} L^{12})a_t . \quad (5.5)$$

After some calculations one obtains

$$y_t = (1 + \theta_1 L + \theta_{s,1} L^{12} + \theta_1 \theta_{s,1} L^{13})a_t \quad (5.6)$$

where  $y_t = (1 - L^{12})(1 - L)x_t$ . Thus, the multiplicative SARIMA model has an ARMA(0,13) representation where only the coefficients

$$\theta_1 , \quad \theta_{12} \stackrel{\text{def}}{=} \theta_{s,1} \quad \text{and} \quad \theta_{13} \stackrel{\text{def}}{=} \theta_1 \theta_{s,1}$$

are not zero. All other coefficients of the MA polynomial are zero.

Thus, we are back in the well-known ARIMA( $p, d, q$ ) world. However, if we know that the original model is a SARIMA(0,1,1)  $\times$  (12,0,1,1), we have to estimate only the two coefficients  $\theta_1$  and  $\theta_{s,1}$ . For the ARMA(0,13) we would estimate instead the three coefficients  $\theta_1$ ,  $\theta_{12}$ , and  $\theta_{13}$ . Thus it is obvious that SARIMA models allow for a parsimonious model building.

In the following, a model specification like (5.6) is called an *expanded model*. In Section 5.4 it is shown, that this kind of specification is required for estimation purposes, since a multiplicative model like (5.5) cannot be estimated directly.

---